Robust tests for one or more allometric lines

S. Taskinen[a,b,*], D.I. Warton[a]

[a] *School of Mathematics and Statistics and Evolution & Ecology Research Centre, The University of New South Wales, NSW 2052, Australia*
[b] *Department of Mathematics and Statistics, FI-40014 University of Jyväskylä, Finland*

Abstract

In allometry, the study of how size variables scale against each other, it is often of interest to fit lines to bivariate data and test hypotheses about slope and elevation about one or several lines. The nature of the problem suggests that bivariate techniques related to principal component analysis are more appropriate than linear regression. Inference methods have been developed for this problem and are in widespread use, however, we demonstrate that such methods are not robust to bivariate contamination, and propose alternative approaches which are. The new approaches use Huber's M-estimator via a plug-in approach, where robust test procedures have the same form as classical ones, but where we plug in robust estimators of parameters and standard errors in place of classical estimators. Simulations demonstrate that these new procedures are robust against bivariate contamination, and can make accurate inferences even from small samples.

Keywords: Analysis of covariance, common slope tests, Huber's M-estimator, major axis, robust statistics, standardised major axis.

1. Introduction

As a field, allometry is sufficiently advanced that textbooks have been written on the subject (Reiss, 1989; Niklas, 1994). Recent examples of allometric research include exploring how leaf venation patterns change with leaf size (Price et al., 2012), and studying how bone tissue and geometry of felid skulls relates to biting force (Chamoli and Wroe, 2011).

It is commonly the case that the relationship between two size variables, possibly after transformation, approximates a straight line. However often the purpose of estimating such allometric lines is not prediction of one variable from the other, rather it is to summarize the relationship between the

*Corresponding author. Tel.: +61 2 9385 7029. E-mail adresses: sara.l.taskinen@jyu.fi (S. Taskinen), david.warton@unsw.edu.au (D. Warton)

two variables in a manner that is symmetric with respect to the X and Y variables (Smith, 2009). As such techniques related to principal components analysis are commonly recommended in preference to linear regression (e.g. Warton et al., 2006), most commonly the major axis (MA) and standardised (or reduced) major axis (SMA), which are respectively the first principal component vector of the variance matrix and of the correlation matrix (rescaled to the original axes), fitted through the centroid of the data. While these methods have occasionally been motivated via errors-in-variables arguments (e.g. McArdle, 2003), it is helpful to regard these methods as a form of data reduction (reducing two dimensions to one), an application for which principal components analysis has long been used.

A number of extensions to the classical approach to allometric line-fitting have been suggested over the years, including amongst others bent-cable regression (Chappell, 1989), non-linear fitting (Ebert and Russell, 1994), methods of comparing several lines (Warton and Weber, 2002; Warton et al., 2006), and Bayesian hierarchical approaches (Price et al., 2009).

One issue that has not been addressed adequately in the literature is robustness of allometric line-fitting methods against outlying observations. Outliers sometimes arise as data entry errors or through measurement error, and these can be relatively easily handled (via correction or repeated measurement). However, it is not unusual for outliers to arise in a more biologically meaningful way – representing observations that should be included in the analyses, while not exerting undue influence (Hampel et al., 1986). An example studied in Taskinen and Warton (2011) was the allometry of offspring mass and body size of mammals – bears were outliers in that they were both unusually large in size and had unusually small offspring per unit body mass. This tendency to be unusual in both the X and Y directions we refer to as *bivariate contamination.* Another example (Fig. 1a, left-most observation) is a plant that is an outlier in the sense that it has both short-lived leaves and unusually small leaf mass per area.

Popular approaches to allometric line-fitting are based on the sample covariance matrix, and as such they lack robustness to outliers (Taskinen and Warton, 2011), potentially leading not only to inefficiency but at times to failure of methods of inference. Classical allometric line-fitting methods have been shown to be robust to non-normality of errors from the line (Warton, 2007), but not to bivariate contamination, under which point estimators can be inefficient and interval estimators inconsistent (Taskinen and Warton, 2011).

In this paper, we will develop robust hypothesis testing approaches for three contexts, illustrated in Fig. 1. In the example of Fig. 1 we consider how leaf lifespan (or leaf longevity) relates to leaf mass per area (LMA), for plant species sampled at two different sites (Wright and Westoby, 2002). This relationship is an example of the "leaf economics spectrum" (Wright et al., 2004), where leaves with higher LMA are more expensive for the plant to construct, but they tend to live longer and hence yield a higher lifetime return.

One problem of interest in this paper is one-sample tests of the slope of an allometric line. In Fig. 1a we are interested in whether leaf longevity is directly proportional to leaf mass per area (a slope of one on log-transformed variables), which would imply no lifetime advantage (in light capture per unit mass of leaf investment) at either end of the leaf economics spectrum.

A second type of inference method of interest in this paper is multi-sample tests of slope – or testing for common slope across several allometric lines. If the slope of the relationship between leaf longevity and LMA changes across different sites (Fig. 1b), then across sites, there are different implications for lifetime gains when moving along the leaf economics spectrum.

A third type of inference method of interest in this paper is multi-sample tests of elevation – is there a shift in elevation across different sites (Fig. 1c)? If so, then leaves in different communities have different opportunities for total lifetime light capture.

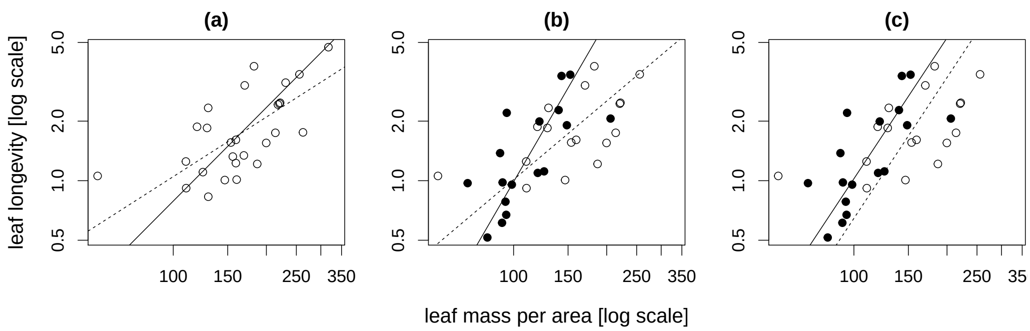

Figure 1: An illustration of the different types of tests considered in this paper using data from Wright and Westoby (2002). Each point represents a different plant species, and points are classified according to whether they were sampled at a site with low soil nutrients (open circles) or high soil nutrients (closed circles). Dashed and solid lines are classical major axis lines computed using samples from these two sites, respectively. We consider (a) one-sample testing of the slope (a slope of one is indicated with a dash-dotted line), (b) common slope testing and (c) testing for common elevation.

"Classical" hypothesis testing methods for each of the three problems, based on the usual sample mean and sample covariance estimators, are reviewed in Warton et al. (2006). Robust alternatives are scarce for these three problems, although some relevant work is available in the literature that will be built upon in this paper.

The problem of robust major axis slope estimation and inference has been considered indirectly in both the principal components (Devlin et al., 1981; Croux and Haesbroeck, 2000) and the errors-in-variables literatures (Zamar, 1989; Cheng and Van Ness, 1992; Fekri and Ruiz-Gazen, 2004). Taskinen and Warton (2011) proposed robust estimators and confidence intervals for slopes of allometric lines, including standardised major axis and major axis. All of the above research focussed on point and interval estimation for the slope coefficient of a single line – none addressed the important problems of deriving robust hypothesis testing procedures for a single allometric line, or for comparing several allometric lines, as depicted in Fig 1a-c. One exception is Boente et al. (2009), who derived a common slopes test applicable in the major axis case (Figure 1b), but whose results are not applicable to standardized major axes, nor to other tests under consideration here.

In this paper, we develop robust approaches to testing hypotheses about parameters in an allometric line, and robust approaches to comparing parameters across several allometric lines – methods that can remain valid in the presence of outliers. We evaluate the accuracy of inferences using each approach, as compared to current approaches in the literature. We propose using Huber's M-estimators (Huber, 1981) for estimation and inference about allometric lines in the presence of outliers, and show that when such estimators are plugged into classical test procedures (Warton et al., 2006), we can achieve quite accurate inferences in small samples.

2. Bivariate line-fitting in allometry

Consider a sample $\boldsymbol{x}_i = (x_{1i}, x_{2i})'$, $i = 1, \ldots, n$, from a bivariate distribution F with finite second moments. We assume that each bivariate observation is independent and identically distributed, and has location vector $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and covariance matrix $\boldsymbol{\Sigma}$. In the following we will assume that the $\boldsymbol{x}_i$ are elliptically distributed, that is, the density function is of the form

$$f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-1/2} f_0\left(\|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})\|\right),$$

where $\boldsymbol{\Sigma}$ is a positive definite symmetric 2×2 matrix (PDS(2)) and $f_0(\|\boldsymbol{z}\|)$ is the density of a spherically distributed random variable $\boldsymbol{z}$, *i.e.* $\boldsymbol{z}$ is a variable whose density is a function of its Euclidean norm $\|\boldsymbol{z}\|$ only.

As done elsewhere (Warton and Weber, 2002; Warton et al., 2006), the major axis (MA) and standardised major axis (SMA) lines are defined to have slope $-\infty < \beta < \infty$ such that

$$\boldsymbol{\Lambda} = \boldsymbol{P}(\beta)'\boldsymbol{\Sigma}\boldsymbol{P}(\beta), \tag{1}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with $\lambda_1 > \lambda_2$ as diagonal values, and $\boldsymbol{P}(\beta)$ is defined differently for the major axis and the standardised major axis, as $\boldsymbol{P}_{\mathrm{MA}}(\beta)$ and $\boldsymbol{P}_{\mathrm{SMA}}(\beta)$ below.

For MA, we assume that $\boldsymbol{P}_{\mathrm{MA}}(\beta)$ is orthonormal:

$$\boldsymbol{P}_{\mathrm{MA}}(\beta_{\mathrm{MA}}) = \frac{1}{\sqrt{1+\beta_{\mathrm{MA}}^2}} \begin{pmatrix} 1 & -\beta_{\mathrm{MA}} \\ \beta_{\mathrm{MA}} & 1 \end{pmatrix}. \tag{2}$$

Hence MA applies a rotation to the original data such that transformed variables are uncorrelated, and most of the variation in the data is along the MA. The columns of $\boldsymbol{P}_{\mathrm{MA}}(\beta_{\mathrm{MA}})$ are the two eigenvectors of $\mathbf{\Sigma}$. The MA slope can then be written as

$$\beta_{\mathrm{MA}} = \frac{1}{2\sigma_{12}} \left(\sigma_{22} - \sigma_{11} + \sqrt{(\sigma_{22}-\sigma_{11})^2 + 4\sigma_{12}} \right), \tag{3}$$

where $\sigma_{12} \neq 0$ and σ_{ij} denotes the (i,j)th element of $\mathbf{\Sigma}$. By definition, the MA slope is rotation equivariant, but not scale equivariant. The axis orthogonal to the major axis is called the minor axis, which has slope $-1/\beta_{\mathrm{MA}}$.

The standardised major axis slope, in contrast, is defined to be scale equivariant, being the major axis fitted to standardised data, then back-transformed to the original scale. Hence, for $\sigma_{12} \neq 0$, the SMA slope β_{SMA} can be shown to satisfy

$$\beta_{\mathrm{SMA}} = sign(\sigma_{12}) \sqrt{\frac{\sigma_{22}}{\sigma_{11}}} \tag{4}$$

and the standardised minor axis slope is $-\beta_{\mathrm{SMA}}$.

The SMA slope β_{SMA} can be written as a solution to the problem given in equation (1) where

$$\begin{aligned} \boldsymbol{P}_{\mathrm{SMA}}(\beta_{\mathrm{SMA}}) &= \frac{1}{\sqrt{|2\beta_{\mathrm{SMA}}|}} \begin{pmatrix} \beta_{\mathrm{SMA}} & -\beta_{\mathrm{SMA}} \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{|\beta_{\mathrm{SMA}}|} & 0 \\ 0 & \frac{1}{\sqrt{|\beta_{\mathrm{SMA}}|}} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}. \end{aligned} \tag{5}$$

Equation (5) re-expresses $\boldsymbol{P}_{\mathrm{SMA}}$ to emphasise the rescaling of data prior to axis fitting. The method of SMA estimation is especially common in biology at the moment, citation statistics suggesting it is used in at least 100 papers per year – the main attraction being its ability to fit a scale equivariant line which is also symmetric (in the sense that the slope of the SMA axis of x_{1i} against x_{2i} is $1/\beta_{\mathrm{SMA}}$, Smith, 2009). For both MA and SMA lines, the elevation is defined so that the axis passes through the center, that is,

$$\alpha_{\mathrm{(S)MA}} = \mu_2 - \beta_{\mathrm{(S)MA}}\mu_1. \tag{6}$$

2.1. Classical estimation of slope and elevation

The MA and SMA slope and elevation estimators are obtained by replacing $\mathbf{\Sigma}$ in (1) with the usual sample covariance matrix $\hat{\mathbf{\Sigma}}_S$ and $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ in (6) by the vector of marginal means $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2)'$. We denote the corresponding slope estimators as $\hat{\beta}_{\text{MA},S}$ and $\hat{\beta}_{\text{SMA},S}$, and elevation estimators as $\hat{\alpha}_{\text{MA},S}$ and $\hat{\alpha}_{\text{SMA},S}$. These estimates can all be derived via maximum likelihood, assuming that the $\boldsymbol{x}_i$ are bivariate normal (Warton et al., 2006). We refer to this throughout as the "classical" approach to estimation, in the sense that estimation makes use of classical estimators of the sample mean and covariance matrix.

2.2. Robust estimation of slope and elevation

The classical (standardised) major axis slope and elevation estimators lack robustness to outliers (Taskinen and Warton, 2011) as they are based on the sample mean and the sample covariance matrix. Robust estimators for MA and SMA slopes and elevations are simply obtained by replacing $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ in (1) and (6) with *any* affine equivariant robust location vector and covariance matrix estimators $\boldsymbol{\mu}_C$ and $\mathbf{\Sigma}_C$ (Fekri and Ruiz-Gazen, 2004; Taskinen and Warton, 2011). By affine equivariance we mean that for any nonsingular 2×2 matrix $\boldsymbol{A}$ and bivariate vector $\boldsymbol{b}$, $(\boldsymbol{A}\boldsymbol{\mu}_C + \boldsymbol{b}, \boldsymbol{A}\mathbf{\Sigma}_C\boldsymbol{A}')$ is the estimator of $(\boldsymbol{\mu}, \mathbf{\Sigma})$ based on $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ whenever $(\boldsymbol{\mu}_C, \mathbf{\Sigma}_C)$ is the corresponding estimator based on $\boldsymbol{x}$.

Many choices of robust estimators $\hat{\boldsymbol{\mu}}_C$ and $\hat{\mathbf{\Sigma}}_C$ are available. As in Taskinen and Warton (2011), we will use Huber's M-estimators (Huber, 1981) in applications because in the bivariate setting this approach is simple, efficient and robust. Huber's M-estimators of the location vector and covariance matrix are implicitly defined as

$$\begin{aligned} \hat{\boldsymbol{\mu}}_M &= [\sum_i w_1(z_i)]^{-1} [\sum_i w_1(z_i) \boldsymbol{x}_i] \\ \hat{\mathbf{\Sigma}}_M &= n^{-1} \sum_i w_2(z_i)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_M)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_M)', \end{aligned} \tag{7}$$

where $z_i = \|\hat{\mathbf{\Sigma}}_M^{-1/2}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_M)\|$ and the weight functions are given by

$$w_1(z) = \min(c/z, 1) \quad \text{and} \quad w_2(z) = \sigma^2 \min(c^2/z^2, 1). \tag{8}$$

In the bivariate case, the cut-off point c is determined by a tuning parameter $0 \le q \le 1$ via the relation $q = F_{\chi^2_2}(c^2)$, and the scaling factor σ^2 is such that $\sigma^2 = F_{\chi^2_4}(c^2) + c^2(1-q)/2$. This choice of σ^2 ensures consistency of the estimator of $\mathbf{\Sigma}$ for multivariate normal data (Maronna, 1976). The influence functions of Huber's M-estimators are derived in Huber (1981) and reviewed in the Appendix. The asymptotic breakdown point is at most 1/3 (Maronna, 1976). This value depends on the tuning parameter q and is obtained by selecting $q = F_{\chi^2_2}(3)$.

3. Robust hypothesis testing for allometric lines

In this section we will review classical approaches to testing hypotheses about slope and elevation, in the one-sample and multi-sample cases, and propose robust alternatives. The classical approaches were reviewed in Warton et al. (2006) and verified to perform quite well in small samples when assumptions are satisfied (Warton et al., 2006). The question of their performance when assumptions are not satisfied will be considered in section 4.

3.1. One-sample tests for slope

Consider a test of whether the true major axis slope is equal to some value b, that is, $H_0 : \beta_{\mathrm{MA}} = b$ or $H_0 : \beta_{\mathrm{SMA}} = b$. In the classical case, "exact" one-sample tests of H_0 can be derived under the assumption that observations $\boldsymbol{x}_i$ are bivariate normal. The approach is well-known, and the derivation in the major axis case is generally attributed to Creasy (1957), and in the standardised major axis case to Pitman (1939).

The major and minor axis scores (f_i and r_i respectively) satisfy

$$(f_i, r_i)' = \boldsymbol{P}'_{\mathrm{MA}}(b)\, \boldsymbol{x_i}. \tag{9}$$

Standardised major and minor axis scores are defined similarly, via $\boldsymbol{P}'_{\mathrm{SMA}}(b)$. Under H_0, from equation (1), (standardised) major and minor axes will be uncorrelated. The classical method of one-sample inference about slope exploits this lack of correlation – we test H_0 by testing for independence of f_i and r_i using classical approaches.

It is well known that when testing for independence between f_i and r_i, $i = 1, \ldots, n$, one may use a linear regression formulation and base the test statistic on $\hat{\beta}_{r|f,S}$, the linear regression coefficient when r_i is regressed against f_i. The resulting test statistic is then given by

$$T = \frac{\sqrt{n-2}\,\hat{\beta}_{r|f,S}}{\sqrt{\widehat{Var}(\hat{\beta}_{r|f,S})}},$$

and under the null hypothesis, $T \sim t_{n-2}$ exactly when data are bivariate normal (e.g. Anderson, 2003). It can be shown that a corresponding two-sided test T^2 can be written as

$$F = \frac{(n-2)R^2_{rf,S}(b)}{1 - R^2_{rf,S}(b)}, \tag{10}$$

where $R_{rf,S}(b)$ denotes the sample correlation coefficient between r_i and f_i. Under the null hypothesis, $F \sim F_{1,n-2}$ exactly for bivariate normal data.

Let us next consider a robust competitor to the classical F test under the assumption of elliptically distributed data. We show in the Appendix that such a test can be derived using the linear regression formulation as described above. This yields to following test statistic

$$X_C^2 = \frac{n\, R_{rf,C}^2(b)}{\hat{\tau}_C(1 - R_{rf,C}^2(b))}, \tag{11}$$

where $R_{rf,C}(b)$ denotes the correlation coefficient between r_i and f_i computed using the robust covariance matrix $\hat{\mathbf{\Sigma}}_C$, and $\hat{\tau}_C$ is an estimate of $\tau_C = 8^{-1}\mathrm{E}_F[\gamma_C^2(z)]$, with $z = \|\mathbf{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})\|$ and γ_C is a weight function in the influence function for $\mathbf{\Sigma}_C$, as in Taskinen and Warton (2011). The term τ_C can be understood as a correction factor for violations of bivariate normality and for use of a different covariance estimator to the classical one. If classical estimators are used, we write this correction factor as τ_S, and note that it equals one exactly when data are bivariate normal, and that it tends to be larger for longer-tailed distributions.

The following is proved in the Appendix.

Theorem 1. *Assume that data are elliptically distributed. Then under the null hypothesis, as $n \to \infty$, $X_C^2 \to_d \chi_1^2$.*

We also consider a small-sample modification of X_C^2, motivated by the form of equation (10):

$$F_C = \frac{(n-2)R_{rf,C}^2(b)}{\hat{\tau}_C(1 - R_{rf,C}^2(b))}, \tag{12}$$

which under the null hypothesis we compare to a $F_{1,n-2}$ distribution. This test can be understood as a plug-in estimate of the exact test F, plugging robust estimates in for their classical counterparts. The F_C statistic simplifies to F in the case when classical estimates are used and data are known to be bivariate normal (since as before $\tau_S = 1$ under bivariate normality). Also note that under H_0, F_C converges to X_C^2 for large n when data come from any elliptical distribution.

3.2. Tests for common slope

Consider now the multi-sample case, where we have g bivariate samples of elliptically distributed data, $\boldsymbol{x}_{ij}$ where $i = 1, \dots, g$ and $j = 1, \dots, n_i$. Each $\boldsymbol{x}_{ij}$ has density $f(\mathbf{x})$ as given previously, with location vector $\boldsymbol{\mu}_i$ and covariance matrix $\mathbf{\Sigma}_i$. We index the true slope and elevation by group ($\beta_{\mathrm{MA},i}$, *etc*). Often we are interested in testing the hypothesis of common slope, that is, $H_0: \ \beta_{\mathrm{MA},i} = b$ or $H_0: \ \beta_{\mathrm{SMA},i} = b$ for all i and for some unknown b. The null hypothesis of common slope can equivalently be written as

$$H_0: \ \boldsymbol{\Sigma}_i = \boldsymbol{P}(b)\boldsymbol{\Lambda}_i\boldsymbol{P}(b)',$$

where $\boldsymbol{P}(b)$ is defined in the major axis case in (2) and in the standardised major axis case in (5).

In the classical case, likelihood ratio tests have been developed for common slope testing (Warton and Weber, 2002) under the assumption of bivariate normality. For major axis estimation, the common slope test is a special case of common principal components analysis that was introduced in Flury (1984). The extension of Flury's likelihood ratio test to heterokurtic case was given recently by Hallin et al. (2010).

Maximum likelihood results due to Flury (1984), when applied in the bivariate case, suggest as a classical common major axis slope estimator $\hat{\beta}_{\mathrm{com},S}$, the solution to:

$$\hat{\boldsymbol{P}}'_{\mathrm{MA},1}\left(\sum_{i=1}^{g} n_i \frac{\hat{\lambda}_{i,1}-\hat{\lambda}_{i,2}}{\hat{\lambda}_{i,1}\hat{\lambda}_{i,2}}\,\hat{\boldsymbol{\Sigma}}_{S,i}\right)\hat{\boldsymbol{P}}_{\mathrm{MA},2} = 0, \tag{13}$$

where we denote by $\hat{\boldsymbol{P}}_{\mathrm{MA},j}$ the jth column of $\boldsymbol{P}_{\mathrm{MA}}(\hat{\beta}_{\mathrm{com},S})$ and $\hat{\lambda}_{i,j}$ is the jth diagonal element of $\hat{\boldsymbol{\Lambda}}_i = diag(\boldsymbol{P}_{\mathrm{MA}}(\hat{\beta}_{\mathrm{com},S})'\hat{\boldsymbol{\Sigma}}_{S,i}\boldsymbol{P}_{\mathrm{MA}}(\hat{\beta}_{\mathrm{com},S})))$. Further, a likelihood ratio test of H_0 reduces to

$$-2\log\hat{\boldsymbol{\Lambda}} = -\sum_{i=1}^{g} n_i \log\left(\frac{|\hat{\Sigma}_{S,i}|}{\hat{\lambda}_{i,1}\hat{\lambda}_{i,2}}\right), \tag{14}$$

and under the null hypothesis, $-2\log\hat{\boldsymbol{\Lambda}} \to_d \chi^2_{g-1}$.

Warton and Weber (2002) extended the approach used in Flury (1984) to the standardised major axis case. The maximum likelihood estimator of $\hat{\beta}_{\mathrm{com},S}$ then solves

$$\hat{\boldsymbol{P}}'_{\mathrm{SMA},1}\left(\sum_{i=1}^{g} n_i \frac{\hat{\lambda}_{i,1}+\hat{\lambda}_{i,2}}{\hat{\lambda}_{i,1}\hat{\lambda}_{i,2}}\,\hat{\boldsymbol{\Sigma}}_{S,i}\right)\hat{\boldsymbol{P}}_{\mathrm{SMA},2} = 0, \tag{15}$$

where $\hat{\boldsymbol{P}}_{\mathrm{SMA},j}$ is now the jth column of $\boldsymbol{P}_{\mathrm{SMA}}(\hat{\beta}_{\mathrm{com},S})$. The likelihood ratio test of the null hypothesis of common slope then equals to that given in (14). Warton and Weber (2002) showed that Bartlett-type adjustment for the likelihood ratio test ensures good performance in small sample sizes, which results in replacement of n_i with $(n_i - 2.5)$ in (14). They also noticed that for both the major and standardised major axis cases, using the affine equivariant properties of $\hat{\Sigma}_{S,i}$, the (Bartlett-corrected) likelihood ratio test statistic may equivalently be written as

$$-2\log\hat{\mathbf{\Lambda}} = \sum_{i=1}^{g}(n_i - 2.5)\,\log\left(1 + \frac{R^2_{rf,i}(\hat{\beta}_{\text{com},S})}{1 - R^2_{rf,i}(\hat{\beta}_{\text{com},S})}\right), \tag{16}$$

where $R_{rf,i}(\hat{\beta}_{\text{com},S})$ is again the correlation coefficient between scores in group i constructed along the (standardised) major and minor axes under common slope $\hat{\beta}_{\text{com},S}$. Hence we have a connection between the form of the multi-sample test of slope and the one-sample test of slope in equation (10).

Connection between the one-sample test and the multi-sample test of slope motivates our approach to robustification of the latter. Robust estimates and tests for common slope are simply obtained by plugging the robust covariance matrix estimates $\hat{\mathbf{\Sigma}}_{C,i}$ into the estimating equations (13) and (15), and the likelihood ratio test statistic (16). The resulting test statistic is then given by

$$X^2_{\text{com},C} = \sum_{i=1}^{g}(n_i - 2.5)\,\log\left(1 + \frac{R^2_{rf,C,i}(\hat{\beta}_{\text{com},C})}{\hat{\tau}_{C,i}(1 - R^2_{rf,C,i}(\hat{\beta}_{\text{com},C}))}\right), \tag{17}$$

where for the ith group $R_{rf,C,i}(\hat{\beta}_{\text{com},C})$ is the correlation coefficient based on the robust covariance matrix estimate $\hat{\mathbf{\Sigma}}_{C,i}$, and $\hat{\tau}_{C,i}$ is the estimate of $\tau_{C,i} = 8^{-1}E[\gamma^2_{C_i}(z)]$. The following theorem gives the null hypothesis distribution of our test statistic.

Theorem 2. *Assume that the data consists of g bivariate samples of elliptically distributed data. Then under H_0, as each $n_i \to \infty$, $X^2_{com,C} \to_d \chi^2_{g-1}$.*

Robust tests for common principal components model were also considered recently in Boente et al. (2009). When testing the null hypothesis of common slope they proceeded as we did above, that is, they replaced the sample covariance matrix in (14) with some robust covariance matrix estimate. The resulting test statistic is

$$X^2_{\text{com},B} = \frac{1}{\tau_C}\sum_{i=1}^{g} n_i \log\left(1 + \frac{R^2_{rf,C,i}(\hat{\beta}_{\text{com},C})}{1 - R^2_{rf,C,i}(\hat{\beta}_{\text{com},C})}\right). \tag{18}$$

The approach used in Boente et al. (2009) differs from ours primarily in estimating a common correction factor τ_C across all samples.

3.3. Robust one-sample test for elevation

Wald tests are typically used to test hypotheses about elevation, whether one-sample or multi-sample, hence we will review asymptotic results for elevation of allometric lines here.

In the Appendix, we use an influence approach (similar to Fekri and Ruiz-Gazen, 2004; Taskinen and Warton, 2011) to show that the limiting distributions of $\sqrt{n}(\hat{\alpha}_{\mathrm{MA},C} - \alpha_{\mathrm{MA}})$ and $\sqrt{n}(\hat{\alpha}_{\mathrm{SMA},C} - \alpha_{\mathrm{SMA}})$ can be derived as normal with mean zero and variances

$$\begin{aligned} \mathrm{ASV}(\hat{\alpha}_{\mathrm{MA},C}) &= \tau_\mu \, \lambda_2 \, (1+\beta^2_{\mathrm{MA}}) + \mu_1^2 \, \mathrm{ASV}(\hat{\beta}_{\mathrm{MA},C}), \\ \mathrm{ASV}(\hat{\alpha}_{\mathrm{SMA},C}) &= \tau_\mu \, \lambda_2 \, 2\beta_{\mathrm{SMA}} + \mu_1^2 \, \mathrm{ASV}(\hat{\beta}_{\mathrm{SMA},C}), \end{aligned} \tag{19}$$

where $\tau_\mu = 2^{-1}\mathrm{E}_F[\gamma_\mu^2(z)]$, z is as before and γ_μ is a weight function in the influence function for $\boldsymbol{\mu}_C$. The asymptotic variance of the (standardised) major axis slope was derived in Taskinen and Warton (2011) as:

$$\begin{aligned} \mathrm{ASV}(\hat{\beta}_{\mathrm{MA},C}) &= \tau_C \, \frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_2)^2}(1+\beta^2_{\mathrm{MA}})^2 \\ \mathrm{ASV}(\hat{\beta}_{\mathrm{SMA},C}) &= \tau_C \, \frac{\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2} \, 4\beta^2_{\mathrm{SMA}}. \end{aligned} \tag{20}$$

Notice that the above results allow us to compute robust large-sample $100(1-\alpha)\%$ confidence intervals for slope (Taskinen and Warton, 2011) and elevation parameters. However such an approach is better suited to inference about elevation because slope estimators tend to converge slowly to a normal distribution (Taskinen and Warton, 2011), being a function of covariance estimators only.

When testing the null hypothesis that the true (standardised) major axis elevation is equal to some value a, that is, $H_0 : \alpha_{\mathrm{MA}} = a$ or $H_0 : \alpha_{\mathrm{SMA}} = a$, we may use as a test statistic

$$T_C = \frac{\sqrt{n}(\hat{\alpha}_{\mathrm{(S)MA},C} - a)}{\sqrt{\widehat{\mathrm{ASV}}(\hat{\alpha}_{\mathrm{(S)MA},C})}}, \tag{21}$$

where $\widehat{\mathrm{ASV}}(\hat{\alpha}_{\mathrm{(S)MA},C})$ is the empirical version of $\mathrm{ASV}(\hat{\alpha}_{\mathrm{(S)MA},C})$ in (19). Under the null hypothesis, $T_C \to_d N(0,1)$. Notice that the asymptotic variance $\mathrm{ASV}(\hat{\alpha}_{\mathrm{(S)MA},C})$ can be rewritten as

$$\begin{aligned} \mathrm{ASV}(\hat{\alpha}_{\mathrm{(S)MA},C}) &= \tau_\mu \begin{pmatrix} -\beta_{\mathrm{(S)MA}} & 1 \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} -\beta_{\mathrm{(S)MA}} \\ 1 \end{pmatrix} + \mu_1^2 \, \mathrm{ASV}(\hat{\beta}_{\mathrm{(S)MA},C}) \\ &= \tau_\mu (\boldsymbol{\Sigma}_{fr})_{22} + \mu_1^2 \, \mathrm{ASV}(\hat{\beta}_{\mathrm{(S)MA},C}), \end{aligned}$$

where τ_μ is defined below equations (19). The asymptotic variance thus consists of two components. The first one equals the variance of the "residual" scores $r_i = x_{2i} - \beta_{\text{(S)MA}}\, x_{1i}$, and the second one takes into account the uncertainty that relates to slope estimation.

3.4. Robust test for common elevation

As in Section 3.2, consider now g bivariate samples of elliptically distributed data. We collect the g location vectors into a $g \times 2$ matrix $\boldsymbol{M} = \begin{pmatrix} \boldsymbol{\mu}_1 & \boldsymbol{\mu}_2 & \cdots & \boldsymbol{\mu}_g \end{pmatrix}'$, where $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i})'$, and the true (standardised) major axis elevations into a g-vector $\boldsymbol{\alpha}_{\text{(S)MA}} = (\alpha_{\text{(S)MA},i}, \ldots, \alpha_{\text{(S)MA},g})'$.

We use a similar approach to Warton et al. (2006) and derive Wald's test statistic under the ellipticity assumption. Let $\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C}$ be the g-vector of elevations based on robust covariance matrix estimates $\hat{\boldsymbol{\Sigma}}_{C,i}$, and $\hat{\boldsymbol{M}}_C$ the matrix consisting of robust location vectors. Then

$$\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C} = \hat{\boldsymbol{M}}_C \begin{pmatrix} -\hat{\beta}_{\text{com},C} \\ 1 \end{pmatrix},$$

where $\hat{\beta}_{\text{com},C}$ is the common slope estimate based on robust covariance matrices $\hat{\boldsymbol{\Sigma}}_{C,i}$ as described in Section 3.2. The null hypothesis of common elevation can now be written in the form $H_0 : L\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C} = \mathbf{0}$, where $L = [\mathbf{1}_{(g-1)\times 1} | - I_{(g-1)\times(g-1)}]$, and tested using a Wald statistic

$$W_C = (L\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C})'(L\,\widehat{\text{ASC}}(\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C})\,L')^{-1}(L\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C}), \tag{22}$$

where $\widehat{\text{ASC}}(\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C})$ is the empirical version of the asymptotic covariance matrix $\text{ASC}(\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C})$. From the asymptotic normality of $\sqrt{n}\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C}$ it then follows that under the null hypothesis, $W_C \to_d \chi^2_{g-1}$.

The asymptotic covariance matrix $\text{ASC}(\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C})$ may be easily derived (using *e.g.* partial influence function approach) and it simplifies to

$$\begin{aligned}\text{ASC}(\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C}) &= diag\left(\tau_{\mu_i} \begin{pmatrix} -\beta_{\text{com}} & 1 \end{pmatrix} \boldsymbol{\Sigma}_i \begin{pmatrix} -\beta_{\text{com}} \\ 1 \end{pmatrix}\right) + \text{ASV}(\hat{\beta}_{\text{com},C})\boldsymbol{M}_1\boldsymbol{M}_1' \\ &= diag(\tau_{\mu_i}(\boldsymbol{\Sigma}_{fr,i})_{22}) + \text{ASV}(\hat{\beta}_{\text{com},C})\boldsymbol{M}_1\boldsymbol{M}_1',\end{aligned}$$

where $diag(a_i)$ is a diagonal matrix with diagonal elements a_i, $\boldsymbol{M}_1$ denotes the first column of $\boldsymbol{M}$ and $\tau_{\mu_i} = 2^{-1}E[\gamma^2_{\mu_i}(r)]$. Similarly to the one-sample case, the first part in $\text{ASC}(\hat{\boldsymbol{\alpha}}_{\text{(S)MA},C})$ equals the variance of residuals $r_i = x_{i2} - \beta_{\text{com}} x_{i1}$, $i = 1, \ldots, g$. The asymptotic variance of the common slope estimate, $\text{ASV}(\hat{\beta}_{\text{com},C})$, can be derived using the partial influence function approach. In the major axis case

$$\text{ASV}(\hat{\beta}_{\text{com},C}) = \left[\sum_{i=1}^{g} n_i\, \tau_{C,i}\, \frac{(\lambda_{i,1} - \lambda_{i,2})^2}{\lambda_{i,1}\lambda_{i,2}}(1 + \beta_{\text{com}})^2\right] \left[\sum_{i=1}^{g} n_i\, \frac{(\lambda_{i,1} - \lambda_{i,2})^2}{\lambda_{i,1}\lambda_{i,2}}\right]^{-2}$$

(Boente and Orellana, 2001; Boente et al., 2002), and in the standardised major axis case

$$\mathrm{ASV}(\hat{\beta}_{\mathrm{com},C}) = \left[\sum_{i=1}^{g} n_i\, \tau_{C,i}\, \frac{(\lambda_{i,1}+\lambda_{i,2})^2}{\lambda_{i,1}\lambda_{i,2}} 4\beta_{\mathrm{com}}^2\right] \left[\sum_{i=1}^{g} n_i\, \frac{(\lambda_{i,1}+\lambda_{i,2})^2}{\lambda_{i,1}\lambda_{i,2}}\right]^{-2}.$$

Notice that by assuming that the correction factor $\tau_{C,i} = 8^{-1}E[\gamma_{C_i}^2(r)]$ is the same for each group, the following is true

$$[\mathrm{ASV}(\hat{\beta}_{\mathrm{com},C})]^{-1} = \sum_{i=1}^{g} [\mathrm{ASV}(\hat{\beta}_{\mathrm{com},C_i})]^{-1},$$

where we denote by $\mathrm{ASV}(\hat{\beta}_{\mathrm{com},C_i})$ the asymptotic variance of the one-sample estimate of slope as given in (20). A similar result was derived for the classical approach in Warton et al. (2006).

4. Finite-sample behaviour of tests

In this section we use simple simulation studies to compare the Type I errors (at the 5% significance level) of the robust tests derived in previous sections to those of classical tests derived under normality assumptions. We base our robust tests on Huber's M-estimators of the location vector and covariance matrix defined in Section 2.2. As mentioned before, the asymptotic results of Section 3 are valid for a broad class of robust estimators, but we prefer using the Huber's M-estimators due to their simplicity and good robustness properties in bivariate cases.

4.1. Finite-sample behaviour of one-sample tests

When comparing one-sample tests for slope or elevation, bivariate samples of sizes $n = 20, 50, 100, 150$ and 200 were generated from the contaminated bivariate normal distribution, that is, the observations were generated from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with probability $(1-\epsilon)$ and from $N(\boldsymbol{\mu}, c^2\boldsymbol{\Sigma})$ with probability ϵ. When testing for slope, the location vector $\boldsymbol{\mu}$ was set to zero, whereas when testing for elevation, the value of $\boldsymbol{\mu}$ was varied. The level of contamination was varied by selecting ϵ and c as follows

(a) bivariate normal model: $\epsilon = 0$,

(b) moderately contaminated model: $\epsilon = 0.1$ and $c = \sqrt{3}$,

(c) strongly contaminated model: $\epsilon = 0.1$ and $c = 3$.

The matrix $\mathbf{\Sigma}$ was generated according to the major axis model in equation (2) or the standardised major axis model in equation (5). In both cases $\beta_{(S)MA} = 1$ and the matrix $\mathbf{\Lambda}$ was chosen to be $\mathbf{\Lambda} = diag(1.5, 0.5)$. The number of replications in all simulation studies was set to 20000, and in all cases, we estimated Type I error at the 0.05 significance level.

Consider first one-sample tests for slope. The null hypothesis to be tested was $H_0 : \beta_{\mathrm{MA}} = 1$ or $H_0 : \beta_{\mathrm{SMA}} = 1$, and the three test statistics to be compared were the: (i) classical F test given in (10); (ii) robust X_C^2 test given in (11); (iii) small-sample corrected robust F_C test given in (12). The simulation results in both the major and the standardised major axis case are given in Table 1.

As seen in Table 1, in the bivariate normal case, the classical F test statistic maintained close to nominal significance level and was more accurate than the two robust tests. For strongly contaminated data, Type I errors of F test were over four times the designated level, even in large samples. As expected, robust X_C^2 test worked well for large sample sizes whereas small-sample corrected F_C test maintained close to nominal significance level also when very small sample sizes were encountered.

Table 1: Type I errors (as a percentage, at the 0.05 level) of the classical F, robust X_C^2 and robust F_C one-sample tests of slope. The errors were computed for data from the (a) bivariate normal distribution, (b) contaminated normal distribution with $\epsilon = 0.1$ and $c = \sqrt{3}$ and (c) contaminated normal distribution with $\epsilon = 0.1$ and $c = 3$. In all cases $\mathbf{\Lambda} = diag(1.5, 0.5)$ and $\beta_{(\mathrm{S})\mathrm{MA}} = 1$.

		(a)			(b)			(c)		
	n	F	X^2	F_C	F	X^2	F_C	F	X^2	F_C
MA	20	5.1	9.6	6.3	6.7	9.6	6.3	16.7	9.5	6.3
	50	5.0	6.7	5.5	7.4	6.6	5.5	20.6	6.6	5.5
	100	4.9	5.8	5.2	7.7	5.9	5.3	22.4	5.8	5.3
	150	5.0	5.4	5.1	7.9	5.5	5.2	22.8	5.5	5.2
	200	4.9	5.3	5.0	7.8	5.4	5.1	23.0	5.4	5.1
SMA	20	4.9	9.7	6.3	6.9	9.8	6.4	16.7	9.8	6.5
	50	5.0	6.6	5.7	7.4	6.7	5.5	20.6	6.7	5.6
	100	5.1	5.7	5.2	7.8	5.9	5.3	22.2	5.6	5.1
	150	4.9	5.3	5.3	7.7	5.5	5.2	23.1	5.5	5.1
	200	4.9	5.1	5.1	8.0	5.3	5.1	23.6	5.4	5.1

Consider next one-sample tests for elevation. We generated our data sets using similar bivariate models as in the previous simulation study. The only difference was that the location vector was set to $\boldsymbol{\mu} = (2, 2)'$, and the null hypothesis to be tested was thus $H_0 : \alpha_{\mathrm{MA}} = 0$ or $H_0 : \alpha_{\mathrm{SMA}} = 0$. The robust

T_C test given in (21) was compared to the classical T test

$$T = \frac{\hat{\alpha}_{(\mathrm{S})\mathrm{MA},S} - a}{\sqrt{\widehat{ASV}(\hat{\alpha}_{(\mathrm{S})\mathrm{MA},S})}},$$

which is derived under the assumption of bivariate normality. Under the null hypothesis, $T \sim t_{n-2}$ approximately (Warton et al., 2006).

Table 2: Type I errors (as a percentage, at the 0.05 level) of the classical T and robust T_C one-sample tests of elevation. The errors were computed for data from the (a) bivariate normal distribution, (b) contaminated normal distribution with $\epsilon = 0.1$ and $c = \sqrt{3}$ and (c) contaminated normal distribution with $\epsilon = 0.1$ and $c = 3$. In all cases $\boldsymbol{\mu} = (2, 2)'$, $\boldsymbol{\Lambda} = diag(1.5, 0.5)$ and $\beta_{(\mathrm{S})\mathrm{MA}} = 1$.

		(a)		(b)		(c)	
	n	T	T_C	T	T_C	T	T_C
MA	20	5.4	9.1	8.1	9.2	15.6	9.2
	50	5.4	6.5	7.1	6.3	17.4	6.5
	100	5.1	5.6	7.3	5.7	19.6	5.7
	150	5.0	5.3	7.4	5.3	20.6	5.3
	200	5.1	5.2	7.5	5.3	21.1	5.3
SMA	20	5.9	8.9	7.6	9.2	17.0	9.5
	50	5.0	5.7	6.6	5.8	17.2	6.1
	100	5.0	5.4	6.9	5.3	18.7	5.5
	150	4.9	5.0	7.0	5.3	19.0	5.3
	200	5.0	5.1	6.9	5.0	19.0	5.1

The simulation results are presented in Table 2. Again, the classical test statistic maintained close to nominal significance levels only in the bivariate normal case. Robust T_C test converged quickly to the designated level, but with very small sample sizes, this test failed to maintain the nominal significance level.

4.2. Finite-sample behaviour of several-sample tests

When comparing tests for common slope and elevation, two independent bivariate samples of sizes $n = 20, 50, 100, 150$ and 200 were generated from (a) bivariate normal model, (b) moderately contaminated model and (c) strongly contaminated model as described in Section 4.1.

When testing for common slope, the null hypothesis of interest was $H_0 : \beta_{\mathrm{MA},i} = b$ or $H_0 : \beta_{\mathrm{SMA},i} = b$ for $i = 1, 2$ and for some unknown b. Three test statistics were included in comparisons: (i) the likelihood ratio test $-2 \log \hat{\Lambda}$

given in (16) and derived under bivariate normality, (ii) robust $X^2_{\text{com},C}$ test given in (17) and (iii) Boente's robust $X^2_{\text{com},B}$ test given in (18). The simulation results based on 20000 replications are given in major axis and in standardised major axis cases in Table 3.

Table 3: Type I errors (as a percentage, at the 0.05 level) of the classical likelihood ratio test $-2\log\hat{\mathbf{\Lambda}}$ for common slope as well as its robust counterparts denoted by $X^2_{\text{com},C} =: X^2_C$ and $X^2_{\text{com},B} =: X^2_B$. The errors are computed for two independent groups of data generated from the (a) bivariate normal distribution, (b) contaminated normal distribution with $\epsilon = 0.1$ and $c = \sqrt{3}$ and (c) contaminated normal distribution with $\epsilon = 0.1$ and $c = 3$. In all cases $\mathbf{\Lambda} = diag(1.5, 0.5)$ and $\beta_{(\text{S})\text{MA}} = 1$.

		(a)			(b)			(c)		
	n	$-2\log\hat{\mathbf{\Lambda}}$	X^2_C	X^2_B	$-2\log\hat{\mathbf{\Lambda}}$	X^2_C	X^2_B	$-2\log\hat{\mathbf{\Lambda}}$	X^2_C	X^2_B
MA	20	3.7	4.5	5.8	4.8	4.4	5.6	11.7	4.4	5.6
	50	4.5	4.9	5.3	6.7	4.9	5.4	18.4	5.0	5.5
	100	4.8	4.9	5.2	7.4	4.9	5.2	21.3	5.0	5.2
	150	4.8	5.0	5.1	7.6	4.9	5.1	22.0	5.0	5.1
	200	4.9	5.0	5.2	7.8	5.0	5.2	22.8	5.0	5.1
SMA	20	4.9	6.3	8.0	6.9	6.5	8.0	17.0	6.5	8.1
	50	5.1	5.7	6.2	7.5	5.6	6.1	20.9	5.5	6.0
	100	5.0	5.3	5.5	7.6	5.3	5.6	22.4	5.3	5.6
	150	4.9	5.1	5.2	7.8	5.2	5.3	23.1	5.2	5.5
	200	5.1	5.1	5.2	7.6	5.0	5.2	23.5	5.1	5.2

Again the likelihood ratio test $-2\log\hat{\mathbf{\Lambda}}$ utilizing the sample mean vectors and sample covariance matrices worked well only in the bivariate normal case. When outlying observations were encountered, this test was not valid. For very large sample sizes, the Type I errors of the two robust tests, $X^2_{\text{com},C}$ and $X^2_{\text{com},B}$, were almost equivalent. In small sample cases, $X^2_{\text{com},C}$ seemed to slightly outperform $X^2_{\text{com},B}$.

Finally, the performances of the two test statistics for common elevation was compared using the same covariance structures and $\boldsymbol{\mu}_1 = (0,0)'$ as before. On this occasion however the location vector of the second sample was set to $\boldsymbol{\mu}_2 = (2,2)'$ and the null hypothesis to be tested was $H_0: \ \alpha_{\text{MA},i} = a$ or $H_0: \ \alpha_{\text{SMA},i} = a$ for $i = 1,2$ and for some unknown a.

The test statistics to be compared were Wald statistics of the form (22). In the classical Wald test W, sample mean vectors and sample covariance matrices were used to estimate $\boldsymbol{\mu}_i$ and $\mathbf{\Sigma}_i$ (Warton et al., 2006), whereas its robust counterpart W_C used Huber's M-estimators.

Simulation results based on 20000 replications are listed in Table 4. In the major axis case, the classical test performed well only when data were normal and sample sizes were large. When highly contaminated data were encountered, the Type I errors exceeded the designated 5% level. The robust Wald test performed poorly in the case of very small sample sizes, but the Type I errors converged quickly to the designated level. Such undesirable small-sample behaviour was not present when testing for common elevation in the standardised major axis case.

Table 4: Type I errors (as a percentage, at the 0.05 level) of the classical W test and robust W_C test for common elevation. The errors were computed for two independent groups of data generated from the (a) bivariate normal distribution, (b) contaminated normal distribution with $\epsilon = 0.1$ and $c = \sqrt{3}$ and (c) contaminated normal distribution with $\epsilon = 0.1$ and $c = 3$. In all cases $\boldsymbol{\mu}_1 = (0,0)'$, $\boldsymbol{\mu}_2 = (2,2)'$, $\boldsymbol{\Lambda} = diag(1.5, 0.5)$ and $\beta_{\text{(S)MA}} = 1$.

		(a)		(b)		(c)	
	n	W	W_C	W	W_C	W	W_C
MA	20	8.0	9.2	10.0	9.7	19.4	10.4
	50	6.0	6.4	8.0	6.6	20.4	6.8
	100	5.5	5.6	7.5	5.5	20.0	5.9
	150	5.3	5.4	7.3	5.3	20.0	5.3
	200	5.1	5.1	7.5	5.3	20.0	5.3
SMA	20	5.5	6.2	6.7	6.3	12.1	6.7
	50	5.1	5.3	6.2	5.4	12.9	5.6
	100	5.1	4.9	6.2	5.0	13.2	5.2
	150	5.1	4.9	6.1	4.9	13.7	5.2
	200	5.0	4.9	6.2	5.0	13.6	5.1

5. Examples

In this section we will compare the performances of robust tests proposed in this paper with the classical tests, using the dataset illustrated previously in Figure 1. We will only report the results for the major axis case, as the results in the standardised major axis case were very similar. The data as well as functions for performing these tests are available in the **R**-package **smatr** (Warton et al., 2012).

In the first column of Figure 2, leaf longevity (in years) is plotted against leaf mass area (kg m^{-2}) for plant species sampled at a site with high level of annual rainfall and a low level of soil nutrients. We wish to test whether leaf longevity is directly proportional to leaf mass per area, that is, $H_0 : \beta_{\text{MA}} = 1$ against $H_1 : \beta_{\text{MA}} \neq 1$. In the first column of Figure 2, the classical major

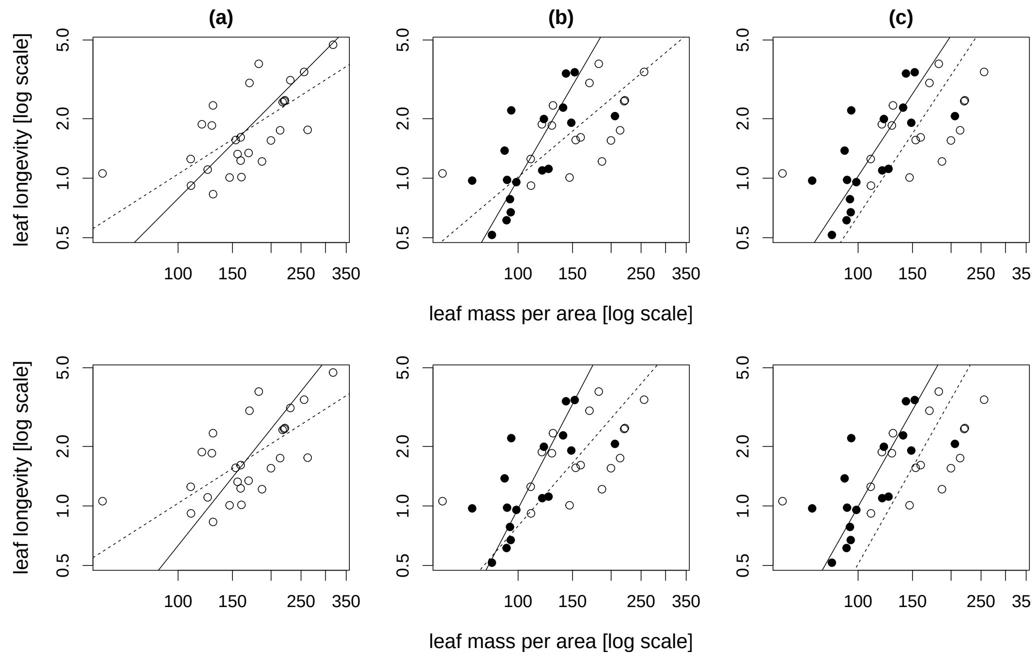

Figure 2: Classical major axis lines (first row) and robust major axis lines based on Huber's M-estimates (second row) as applied to leaf data from Wright and Westoby (2002). Each point represents a different plant species, and points are classified according to whether they were sampled at a site with low soil nutrients (open circles) or high soil nutrients (closed circles). Dashed and solid lines are major axis lines computed using samples from these two sites, respectively. We consider (a) one-sample testing of the slope (a slope of one is indicated with a dash-dotted line), (b) common slope testing and (c) testing for common elevation.

axis line (top row), robust major axis line (bottom row) and a line of slope one (both rows) are plotted. We readily notice that the classical major axis line is affected by one outlier in the lower left corner, and the estimated slope is noticeably flatter than that of the robust line. The classical and robust F test statistics, with corresponding p-values, are listed in Table 5. It is seen that, with robust F_C test, weak evidence against the null hypothesis is detected. Due to the one outlying observation, the classical F test statistic is very small and thus provides no evidence against the null hypothesis.

In the second column of Figure 2, leaf longevity and leaf mass per area are plotted for plant species sampled from high-rainfall sites with low and high nutrient contrasts, respectively. Let us now investigate whether the relationship between variables changes across these two sites. We will thus test for a common major axis slope, that is, the null hypothesis of interest is $H_0 : \beta_{\mathrm{MA},i} = b$ against $H_1 : \beta_{\mathrm{MA},i} \neq b$ for $i = 1, 2$ for some unknown b. As

Table 5: Classical and robust test statistics, with corresponding p-values, when testing for (a) major axis slope equals one, (b) common major slope and (c) common major axis elevation. The data are presented in Figure 2.

	Classical test	Robust test
(a)	$F = 0.609$ $p = 0.447$	$F_C = 3.486$ $p = 0.082$
(b)	$-2 \log \hat{\mathbf{\Lambda}} = 2.003$ $p = 0.157$	$X^2_{\text{com},C} = 1.250$ $p = 0.264$
(c)	$W = 2.919$ $p = 0.088$	$W_C = 4.967$ $p = 0.026$

before, when comparing classical major axis lines given in the first row to the robust major axis lines given in the second row, we see that one outlier present in the low nutrient site affects the corresponding classical major axis line. However, the outlier does not have much influence on the classical likelihood ratio test and, similarly to the robust test, the null hypothesis of interest is accepted.

Finally, let us investigate whether there is a shift in elevation when comparing plant species sampled from two different sites. The null hypothesis to be tested is thus $H_0 : \alpha_{\text{MA},i} = a$ against $H_1 : \alpha_{\text{MA},i} \neq a$ for $i = 1, 2$ and some unknown a. The sites to be compared are again those with low and high nutrient contrasts, respectively. Corresponding data as well as classical and robust major axis lines sharing common slope are illustrated in the third column of Figure 2. The corresponding major axis lines are relatively similar. However, some difference is seen between classical and robust Wald statistics reported in Table 5. The classical Wald test is unable to reject the null hypothesis at the 0.05 significance level and based on that we cannot state that there is a clear shift in elevation across these two sites of interest. The p-value based on the robust test is clearly below 0.05, suggesting that greater lifetime gains are available to leaves at higher nutrient sites – because leaves of a given mass per area tend to be kept on the plant for longer.

6. Discussion

Allometric lines are commonly fitted using major and standardised major axes, and biologists routinely wish to test one-sample hypotheses or multiple-sample hypotheses about slope and elevation. We have demonstrated that the methods currently used (Warton et al., 2006) can be quite sensitive to contamination. Further, we have shown that these methods can be robustified in a relatively simple way. Our simulations (Tables 1-4) indicate that the proposed tests have good performance across a range of conditions. They maintained near-exact coverage for contaminated data, although were slightly liberal in small samples, more so when making inferences about elevation. There is a case for continuing to use classical inference methods when their assumptions are likely to be satisfied, due to more exact Type I error in small samples, but in other settings there is a strong case for using robust methods in their place.

The robust methods derived here relax the assumption of bivariate normality to an assumption that data are elliptically distributed. This latter assumption is more biologically realistic, given that allometric data are often heavy-tailed in distribution (Robinson and Hamann, 2011, for example), while the bivariate normal model is still available as a special case. Another important special type of elliptical distribution is the contaminated normal mixture model, a standard model for capturing outliers (Hampel et al., 1986) such as the left-most observation in Fig. 1a.

Taskinen and Warton (2011) previously considered the problem of interval estimation for the slope of a (standardised) major axis, using Wald-type intervals. This problem is closely related to one-sample testing of slope, given that one could construct a confidence interval via inversion of a one-sample test statistic. Taskinen and Warton (2011) constructed confidence intervals for slope by exploiting the asymptotic normality of $\hat{\beta}_{(\mathrm{S})\mathrm{MA},C}$, but this had poor small-sample performance and required a bootstrap for remediation. However, given the very good performance of F_C in Table 1, it seems one can achieve comparable coverage probabilities to the bootstrap much faster via inversion of F_C and use of critical values from the $F_{1,n-2}$ distribution. This approach to interval estimation is actually used in the classical case (Warton et al., 2006), so what we suggest here is to take the classical approach, known to have good small-sample performance when assumptions are satisfied, then plug-in robust estimates as required, and use the same critical values from small-sample distributions as motivated formally in the classical case.

The most successful method of robustification considered in this paper was the approach just described – to take classical tests known to have good small-sample performance, and replace classical estimators of mean, covariance, and standard error with robust estimators. The plug-in strategy was broadly successful and has the advantages of being simple and applicable whenever small-sample methods are available in the classical case. It would be interesting to see how effectively our robust plug-in approach works in other contexts outside of allometry.

Acknowledgements

The work of Sara Taskinen was supported by the Academy of Finland. David Warton is supported by an Australian Research Council Future Fellowship (project number FT120100501).

References

Anderson, T. W., 2003. An Introduction to Multivariate Statistical Analysis (Third ed.). Wiley, New York.

Boente, G., Orellana, L., 2001. A robust approach to common principal components. In: Fernholz, L., Morgenthaler, S., Stahel, W. (Eds.), Statistics in Genetics and in the Environmental Sciences. Birkhauser, pp. 117–147.

Boente, G., Pires, A. M., Rodrigues, I. M., 2002. Influence functions and outlier detection under the common principal components model: A robust approach. Biometrika 89, 861–875.

Boente, G., Pires, A. M., Rodrigues, I. M., 2009. Robust tests for the common principal components model. Journal of Statistical Planning and Inference 139, 1332–1347.

Chamoli, U., Wroe, S., 2011. Allometry in the distribution of material properties and geometry of the felid skull: Why larger species may need to change and how they may achieve it. Journal of Theoretical Biology 283 (1), 217 – 226.

Chappell, R., 1989. Fitting bent lines to data, with applications to allometry. Journal of Theoretical Biology 138 (2), 235 – 256.

Cheng, C.-L., Van Ness, J. W., 1992. Generalized m-estimators for errors-in-variables regression. Annals of Statistics 20, 385–397.

Creasy, M. A., 1957. Confidence limits for the gradient in the linear functional relationship. Journal of the Royal Statistical Society B 18, 65–69.

Croux, C., Haesbroeck, G., 2000. Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. Biometrika 87, 603–618.

Devlin, S. J., Gnandesikan, R., Kettenring, J. R., 1981. Robust estimation of dispersion matrices and principal components. Journal of the American Statistical Association 76, 354–362.

Ebert, T. A., Russell, M. P., 1994. Allometry and model ii non-linear regression. Journal of Theoretical Biology 168 (4), 367 – 372.

Fekri, M., Ruiz-Gazen, A., 2004. Robust weighted orthogonal regression in the errors-in-variables model. Journal of Multivariate Analysis 88, 89–108.

Flury, B. N., 1984. Common principal components in k groups. Journal of the American Statistical Association 79, 892–898.

Hallin, M., Paindaveine, D., Verdebout, T., 2010. Testing for common principal components under heterokurticity. Journal of Nonparametric Statistics 22, 879–895.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. J., 1986. Robust Statistics: the Approach Based on Influence Functions. Wiley, New York.

Huber, P. J., 1981. Robust Statistics. Wiley, New York.

Maronna, R. A., 1976. Robust m-estimators of multivariate location and scatter. Annals of Statistics 4, 51–67.

McArdle, B. H., 2003. Lines, models, and errors: Regression in the field. Limnology and Oceanography 48, 1363–1366.

Niklas, K. J., 1994. Plant Allometry: The Scaling of Form and Process. University of Chicago Press, Chicago.

Pitman, E. T. G., 1939. A note on normal correlation. Biometrika 31, 9–12.

Price, C. A., Ogle, K., White, E. P., Weitz, J. S., 2009. Evaluating scaling models in biology using hierarchical bayesian approaches. Ecology letters 12 (7), 641–651.

Price, C. A., Wing, S., Weitz, J. S., FEB 2012. Scaling and structure of dicotyledonous leaf venation networks. Ecology Letters 15 (2), 87–95.

Reiss, M. J., 1989. The allometry of growth and reproduction. Cambridge University Press, Cambridge.

Robinson, A. P., Hamann, J. D., 2011. Forest Analytics with R. Springer Science+ Business Media.

Smith, R. J., 2009. Use and misuse of the reduced major axis for line-fitting. American Journal of Physical Anthropology 140, 476–486.

Taskinen, S., Warton, D. I., 2011. Robust estimation and inference for bivariate line-fitting in allometry. Biometrical Journal 53, 652–672.

Warton, D., Duursma, R., Falster, D., Taskinen, S., 2012. smatr 3 - an r package for estimation and inference about allometric lines. Methods in Ecology and Evolution 2, 257–259.

Warton, D. I., 2007. Robustness to failure of assumptions of tests for a common slope amongst several allometric lines – a simulation study. Biometrical Journal 49 (2), 286–299.

Warton, D. I., Weber, N. C., 2002. Common slope tests for errors-in-variables models. Biometrical Journal 44 (2), 161–174.

Warton, D. I., Wright, I. J., Falster, D. S., Westoby, M., 2006. Bivariate line-fitting methods for allometry. Biological Reviews 81 (2), 259–291.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., Midgley, J. J., Navas, M. L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V. I., Roumet, C., Thomas, S. C., Tjoelker, M. G., Veneklaas, E. J., Villar, R., 2004. The worldwide leaf economics spectrum. Nature 428, 821–827.

Wright, I. J., Westoby, M., 2002. Leaves at low versus high rainfall: coordination of structure, lifespan and physiology. New Phytologist 155, 403–416.

Zamar, R., 1989. Robust estimation in the errors-in-variables model. Biometrika 76, 149–160.

Appendix: Proofs of the results

The asymptotic variance of $\sqrt{n}(\hat{\alpha}_{\mathbf{(S)MA},C} - \alpha_{\mathbf{(S)MA}})$. Write first $\boldsymbol{\mu}_C(F)$ and $\boldsymbol{\Sigma}_C(F)$ for the functionals corresponding to any affine equivariant estimators $\boldsymbol{\mu}_C$ and $\boldsymbol{\Sigma}_C$. Corresponding influence functions at spherical F are then given by

$$\mathrm{IF}(\boldsymbol{x}; \boldsymbol{\mu}_C, F) = \gamma_\mu(z)\boldsymbol{u} \quad \text{and} \quad \mathrm{IF}(\boldsymbol{x}; \boldsymbol{\Sigma}_C, F) = \gamma_C(z)\boldsymbol{u}\boldsymbol{u}^T - \delta_C(z)\boldsymbol{I}, \tag{23}$$

where $z = \|\boldsymbol{x}\|$, $\boldsymbol{u} = \|\boldsymbol{x}\|^{-1}\boldsymbol{x}$, and weight functions γ_μ, γ_C and δ_C depend on the distribution F and the functionals $\boldsymbol{\mu}_C(F)$ and $\boldsymbol{\Sigma}_C(F)$, respectively (Hampel et al., 1986). For Huber's M-estimators with weights given in (8), $\gamma_\mu(z) = \eta_1^{-1} w_1(z) z$, where $\eta_1 = 2^{-1} E_F[2w_1(z) + r w_1'(z)]$, and $\gamma_C(z) = \eta_2^{-1} w_2(z) z^2$, where $\eta_2 = 8^{-1} \mathrm{E}_F[4w_2(z)z^2 + w_2'(z)z^3]$. Influence functions in the elliptical case follow from the affine equivariance properties of the estimators.

By (6), the influence function of the robust elevation functional $\alpha_{\mathrm{(S)MA},C}(F)$ at elliptical F with location $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is

$$\mathrm{IF}(\boldsymbol{x}; \alpha_{\mathrm{(S)MA},C}, F) = (-\beta_{\mathrm{(S)MA}}, 1)\,\mathrm{IF}(\boldsymbol{x}; \boldsymbol{\mu}_C, F) - \mu_1\,\mathrm{IF}(\boldsymbol{x}; \beta_{\mathrm{(S)MA},C}, F),$$

where

$$\mathrm{IF}(\boldsymbol{x}; \beta_{\mathrm{MA},C}, F) = \gamma_C(z) \frac{\sqrt{\lambda_1 \lambda_2}}{(\lambda_1 - \lambda_2)} \,(1 + \beta_{\mathrm{MA}}^2)\, u_1 u_2$$
$$\mathrm{IF}(\boldsymbol{x}; \beta_{\mathrm{SMA},C}, F) = \gamma_C(z) \frac{\sqrt{\lambda_1 \lambda_2}}{\lambda_1 + \lambda_2}\, 2\beta_{\mathrm{SMA}}\, u_1 u_2,$$

with $z = \|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})\|$ and $\boldsymbol{u} = z^{-1}(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})) = (u_1, u_2)'$ uniformly distributed on the unit sphere. The asymptotic variance then reduces to

$$\begin{aligned}\mathrm{ASV}(\hat{\alpha}_{\mathrm{(S)MA},C}) &= \mathrm{E}_F[\mathrm{IF}^2(\boldsymbol{x}; \alpha_{\mathrm{(S)MA},C}, F)] \\ &= (-\beta_{\mathrm{(S)MA}}, 1)\,\mathrm{ASV}(\hat{\boldsymbol{\mu}}_C)\,(-\beta_{\mathrm{(S)MA}}, 1)' + \mu_1^2\,\mathrm{ASV}(\hat{\beta}_{\mathrm{(S)MA}}),\end{aligned}$$

as $\mathrm{E}_F[\mathrm{IF}(\boldsymbol{x}; \boldsymbol{\mu}_C, F)\mathrm{IF}(\boldsymbol{x}; \beta_{\mathrm{(S)MA},C}, F)] = \mathbf{0}$ by symmetry. The result then follows as $\mathrm{ASV}(\hat{\boldsymbol{\mu}}_C) = 2^{-1}\mathrm{E}_F[\gamma_\mu^2(z)]\boldsymbol{\Sigma}$.

Proof of Theorem 1. As mentioned in Section 3.1, we may base our test statistic on $\hat{\beta}_{r|f,C}$ which is a robust regression coefficient obtained when r_i is regressed against f_i, that is, $\hat{\beta}_{r|f,C}$ is such that it satisfies

$$\begin{pmatrix} 1 & 0 \\ -\hat{\beta}_{r|f,C} & 1 \end{pmatrix} \hat{\boldsymbol{\Sigma}}_{fr,C} \begin{pmatrix} 1 & -\hat{\beta}_{r|f,C} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \hat{\lambda}_1 & 0 \\ 0 & \hat{\lambda}_2 \end{pmatrix}, \tag{24}$$

where $\hat{\boldsymbol{\Sigma}}_{fr,C}$ is the robust covariance matrix based on the scores f_i and r_i. The limiting distribution of $\hat{\beta}_{r|f,C}$ can be easily derived using the influence function approach and the affine equivariance properties of $\hat{\boldsymbol{\Sigma}}_C$. We then replace $\hat{\beta}_{r|f,C}$ and $\hat{\boldsymbol{\Sigma}}_{fr,C}$ in (24) with corresponding functionals and solve $\beta_{r|f,C}(F)$ as a function of $\boldsymbol{\Sigma}_C(F)$. Then (23) yields to

$$IF(\boldsymbol{x}; \beta_{r|f,C}, F) = \gamma_C(z) \frac{\sqrt{\lambda_1 \lambda_2}}{\lambda_1} u_1 u_2,$$

where γ_C and $\boldsymbol{u} = (u_1, u_2)'$ are as given in Section 2.2. From the asymptotic normality of $\hat{\boldsymbol{\Sigma}}_C$ it then follows that $\sqrt{n}\hat{\beta}_{r|f,C}$ is asymptotically normal with mean zero (under H_0) and its limiting variance can be written as

$$\mathrm{ASV}(\hat{\beta}_{r|f_C}) = \tau_C \, \frac{\lambda_1 \lambda_2}{\lambda_1^2} = \tau_C \, \frac{|\boldsymbol{\Sigma}_{fr,C}|}{(\boldsymbol{\Sigma}_{fr,C})_{11}^2}$$

where $\tau_C = 8^{-1} E[\gamma_C^2(z)]$. By writing $\widehat{\mathrm{ASV}}(\hat{\beta}_{r|f})$ for the empirical version of $\mathrm{ASV}(\hat{\beta}_{r|f})$, we finally notice that our test statistics can be rewritten as

$$X_C^2 = \frac{n \, \hat{\beta}_{r|f,C}^2}{\widehat{\mathrm{ASV}}(\hat{\beta}_{r|f,C})} = \frac{n \, R_{rf,C}^2(b)}{\hat{\tau}_C (1 - R_{rf,C}^2(b))},$$

where $R_{rf,C}(b)$ is the correlation coefficient between r_i and f_i computed using $\hat{\boldsymbol{\Sigma}}_C$, and $\hat{\tau}_C$ is the estimate of τ_C. Therefore, under the null hypothesis $X_C^2 \to_d \chi_1^2$.

Proof of Theorem 2. We will consider the distribution of the common slopes statistic in both the cases when $\beta_{\mathrm{com},C}$ is known and unknown, and write these statistics as $X^2(\beta_{\mathrm{com},C})$ and $X^2(\hat{\beta}_{\mathrm{com},C})$, respectively.

Assume first that $\hat{\beta}_{\mathrm{com},C}$ is fixed, and to shorten notation write $R_i = R_{rf,C,i}(\beta_{\mathrm{com},C})$. Using a Taylor expansion of $\log(1+x)$, $X^2(\beta_{\mathrm{com},C})$ can be written as:

$$X^2(\beta_{\mathrm{com},C}) = \sum_{i=1}^{g} \left(\frac{n_i R_i^2}{\hat{\tau}_{C,i}(1 - R_i^2)} + \frac{1}{2n_i} \left(\frac{n_i R_i^2}{\hat{\tau}_{C,i}(1 - R_i^2)} \right)^2 + \dots \right)$$

Theorem 1 now implies that under H_0, the leading term in the summation of $X^2(\beta_{\mathrm{com},C})$ independently follows a χ_1^2 distribution for each i and the remaining terms are ignorable, being $O_p(n_i^{-1})$ or smaller. Hence $X^2(\beta_{\mathrm{com},C}) \to_d \chi_g^2$.

Further, the difference between the test statistics $X^2(\beta_{\mathrm{com},C}) - X^2(\hat{\beta}_{\mathrm{com},C})$ tests the hypothesis that the common slope equals $\beta_{\mathrm{com},C}$. Following arguments similar to Theorem 1, we can deduce from the asymptotic normality of $\hat{\beta}_{\mathrm{com},C}$ that $X^2(\beta_{\mathrm{com},C}) - X^2(\hat{\beta}_{\mathrm{com},C}) \to_d \chi_1^2$ independently of $X^2(\hat{\beta}_{\mathrm{com},C})$, hence $X^2(\hat{\beta}_{\mathrm{com},C}) \to_d \chi_{g-1}^2$.